Bureau of Rural Science
Department of Primary Industry

LIVESTOCK DISEASE SURVEYS

A FIELD MANUAL FOR VETERINARIANS

R.M. Cannon and R.T. Roe

Australian Government Publishing Service
Canberra 1982

ISBN 0 644 02101 2

First published 1982 for the Australian Bureau of Animal Health
Reprinted 1986 for the Bureau of Rural Science

Printed in Australia by Better Printing Service, 44 Paterson Parade, Queanbeyan, N.S.W.

CONTENTS

mal Health

Parade, Queanbeyan, N.S.W.

INTRODUCTION

The Epidemiology Branch of the Bureau is frequently asked questions such as 'How many animals do we need to test to determine whether a certain disease is present in a herd or flock?' or 'How should we carry out a survey to find the prevalence of a disease in a State?'. This manual is an attempt to make the answers to these and similar questions available to field veterinarians in an easy-to-follow form.

The material included in this manual is not original. All the principles and concepts discussed are covered in texts on sampling theory or statistical methods. However, it is not always easy for a veterinarian to put his hand on an appropriate book when required and the treatment given in such books is often in unfamiliar terms. We have attempted to present the same material in the context of the types of questions we are frequently asked.

Section 1 deals with questions that must be answered when attempting to detect the presence of a disease, while Section 2 deals with estimating the level of occurrence of a disease present in a population. Section 3 considers sampling methodology and Section 4 discusses some commonly used (and misused) terms and concepts. Tables are included to provide the answers to the problems discussed in Sections 1 and 2.

For those who wish to obtain further detail of the principles and mathematical theory behind sampling techniques, a series of Appendices and a list of further reading are included.

1. SURVEYS TO DETECT THE PRESENCE OF DISEASE

1.1 Proving that a Disease is Not Present

Field veterinarians are frequently asked to certify the freedom of animal groups or populations from certain diseases. What testing needs to be undertaken to be able to give such a certification?

In some cases those wanting the certification will prescribe tests that must be conducted on all animals in the group. In this case no sampling considerations arise - all must be tested.

In other cases the veterinarian may be asked to certify as to the disease status of the property of origin of the animals moving interstate or overseas. In most cases the certification required will be that, to the best of the certifying officer's knowledge and belief, there has been no evidence of the disease in question on the property of origin of the stock during the previous 12 months. In such a case, so long as the certifying officer can satisfy himself that the animals in question have not been off the property in the previous 12 months, the certification could be given on the basis of records and knowledge of the disease history of the property.

However, in some cases the veterinarian may be asked to certify that the property of origin is free of a certain disease. In the absence of records of whole herd tests for the specified diseases, this would require the testing of at least a sample of the animals on the property. The sample tested should be of sufficient size to give 95% confidence that the disease is not present at a prevalence that is substantially below the prevalence expected on a diseased property. This question will be answered in section 1.2.

To prove that a disease is not present in a population, particularly if the disease could be present at a low level, requires the testing of all susceptible individuals in the population.

1.2 Detecting the Presence of a Disease

The size of the sample which needs to be tested to determine whether a disease is present in a particular population depends on:

a) the size of the animal population in question

b) if the disease is present, the likely prevalence

c) the reliability required of the conclusions.

a) Size of the population to be sampled

If the sample makes up a significant proportion of the total population, the progressive sampling of negative animals from the population increases the probability that the next animal sampled will be positive. This effect is insignificant if the population to be surveyed is very large compared to the sample size.

b) <u>The expected level of occurrence of the condition, if present</u>

In most cases if a disease is present in a herd or flock we can expect some proportion, rather than a single animal, to be infected. An estimate of the likely level of occurrence can be made from the epidemiology of the disease and results of surveys in other herds or flocks. The higher the prevalence, the smaller the sample size that is required to detect an infected animal.

Thus a sample size can be chosen such that, if the sample is negative, it may be concluded that it is unlikely that any animal in the herd or flock has the disease.

c) <u>Reliability of the conclusions drawn</u>

By adopting a sampling procedure we are accepting, for the sake of economy, that we will not be able to make an absolute statement based on our findings. The larger the sample, the greater the confidence that can be placed in the results. The sample size can be chosen to give the confidence level desired.

The confidence levels that have generally been used in disease surveys are 95% or, occasionally, 99%. For example, suppose that the sample size has been chosen to give us 95% confidence of detecting antibodies if they are present in 10% of the herd. This means that, on average, 5 out of 100 herds with a 10% prevalence would not be detected as infected by our survey. However, very few herds with an antibody prevalence greater than 10% would not be detected by the survey, while the error rate for herds with a prevalence less than 10% would be greater than 5 in 100. Thus, if we expect infected herds to contain 20% to 50% positive animals, the testing of a sample that gave 95% confidence of detecting a 10% level of infected animals would detect most infected herds.

Table 1 gives the sample size required to detect various minimum levels of infection in different sized herds or flocks at one of three confidence levels.

Tables 2 and 3 look at this information in two other ways. Table 2 illustrates the effect that different proportions of sampling have on the chances of detecting a positive in populations with a small number of positives, while Table 3 shows the effect of sample size on the probability of detecting a positive in a population with a specified proportion of positives.

1.3 Disease Monitoring and Disease Surveillance

Disease monitoring poses a different problem. Consider, as an example, abattoir surveillance for brucellosis. We know the proportion of the herd that has been tested at slaughter. If all of these were negative, we wish to know if this is sufficient to say that the herd is negative, and so avoid doing a whole herd test.

The question to be answered is thus the converse of the one faced in Section 1.1. Rather than wanting to determine the sample size required for a given level of confidence of detecting disease, we want an indication of the level of disease that may be present given that our surveillance sample is negative. Table 1 can also be used to answer this question and gives the upper limit for the number of positives that could be present in the population given that none were detected in the surveillance sample. Again Tables 2 and 3 provide an alternate way of looking at the problem.

A Special Word About Abattoir Surveillance for Disease

Tables 1 to 3 assume that a random sample is tested. However, slaughter animals are not a true random sample of the population. Animals sent for slaughter are those judged by their owners to be in a suitable condition. Most of these animals will be free from any clinical disease. Slaughter animals will also be unrepresentative of the age structure and sex ratio of the population from which they are derived. Individuals with a poor performance in such things as reproduction or milk production are likely to be over-represented in slaughter animals.

Care must therefore be exercised in drawing inferences about the prevalence of disease in a population based on abattoir survey results. While information on the prevalence of conditions that do not affect the preparation of animals for slaughter (and do not affect the likelihood of an animal being culled) may be obtained from carefully planned abattoir surveys, thought needs to be given to the possible biases in this method before commencing a survey.

On the other hand, trace-back from abattoir surveillance does provide a cost effective method of identifying diseased herds and flocks. Abattoir surveillance data can also be cumulated from one period to the next - thereby building up a record on the herds or flocks of origin. In this respect the tables are not strictly correct since the animals in the herd or flock for which the percentage monitored is calculated is a changing population with replacements being bred or introduced. The effect of this, however, will generally be to increase the probability of detecting disease above the levels given in Table 1.

2. SURVEYS TO MEASURE THE PREVALENCE OF DISEASE

2.1 Establishing the Level of Occurrence of a Disease

Estimation of the prevalence of a disease involves the same sampling theory as detecting the presence of disease. In fact, surveying for the presence of a disease is a special application of the procedure followed in estimating the prevalence of disease in a population.

Again the sample size required will depend on the level of the disease in the population, the confidence level desired and the size of the population.

However, the question usually posed in relation to surveying to determine the prevalence of disease is, 'Having tested a random sample of animals from a population of known size and found a certain proportion of these to be positive, what does this mean about the level of disease in the population?'.

Table 4 gives an estimate of the number of animals that need to be tested if it is desired to estimate, with a specified confidence level, the proportion within the population that have the disease.

Table 5 gives the possible range for the disease prevalence in a population, for three different confidence levels, that can be deduced from the findings of a survey test of a given number of animals. If more precision is required, then it is necessary to test additional randomly selected animals from the same population until the required precision is obtained.

3. SAMPLING METHODS

In the previous sections concerning the size of the sample required for various purposes, it has been assumed that the sample to be surveyed would be selected at random. Without proper randomisation, we cannot validly extrapolate from the sample we have surveyed to make inferences about the population it is supposed to represent.

Random does not mean haphazard. It has been demonstrated repeatedly that it is virtually impossible to prevent the introduction of some form of bias, either consciously or sub-consciously, in selecting a sample by means other than a randomizing process. To select a proper random sample requires rigorous adherence to a predetermined random selection procedure.

The four common sampling methods are described below. These are often used in various combinations.

3.1 Simple Random Sampling

A simple random sample is a sample selected such that each animal in the population being surveyed has an equal opportunity of being included in the sample. The tables included in this booklet for determining sample size etc. are all based on the selection of a simple random sample. If the sample tested is not a simple random sample these tables do not apply and larger samples will be required.

A random sample may be selected by numbering each animal in the population (such as by ear-tag numbers), writing these numbers on slips of paper, mixing them in a hat, and drawing the required number of slips from the hat to identify the members of the sample. A more convenient way of choosing the numbers is to use a table of random numbers such as Table 6.

3.2 Systematic Sampling

In systematic sampling, animals are selected for inclusion in the sample at equal intervals from the (ordered) population. If a sample of 1/n of the population is required, we would start at a randomly selected animal in the first n animals, and thereafter sample every n^{th} animal.

For a systematic sample, the population size need not be known before the survey starts. This is particularly useful when prospectively sampling events such as live births or laboratory accessions. To select a random sample containing 1/n of a population, the size of the population must be known (at least approximately).

Systematic sampling will usually be much easier than random sampling and on occasions may be the only practicable method. However, care must be taken in systematic sampling, if cyclic fluctuations exist in the population, that the sample interval does not coincide with the fluctuation interval.

Systematic sampling is appropriate for quality control testing on production lines.

3.3 Stratified Sampling

The population of a livestock species on any property is normally managed in a number of discrete herds or flocks. Where the sample to be selected is very small in comparison with the total population, it is quite likely that a simple random sample selected from the total population, could, by chance, severely under-represent (or completely exclude) some herds or flocks. A random sample of <u>herds</u> will under-represent animals in large herds, whereas a random sample of <u>animals</u> will under-represent small herds. This can be avoided by stratified sampling, which ensures that each group or unit in the aggregate population is adequately represented.

For example, suppose we wish to sample 58 sheep from a property running 7800 sheep as part of a survey. (A sample of 58 would give 95% confidence of detecting the presence of a condition if present at a prevalence of 5% or greater.) If the 7800 sheep are run in four separate flocks, we should sample from each flock in the same proportion it is to the total population (rounding up to the next whole number in each case).

	number in flock	number sampled
Ewes	3000	58 x 3000 ÷ 7800 = 23
Maiden ewes	800	6
Weaners	2500	19
Wethers	1500	12
	7800	60

If the animals selected from each flock are a random sample of that flock, we then have a stratified random sample, which can be termed a random sample stratified by flock size of the sheep on that property.

3.4 Cluster Sampling

In cluster sampling, groups or clusters of individuals are chosen for testing. The groups may be chosen at random or systematically, but no control is exercised over membership of each group. Cluster sampling is frequently resorted to where no reliable list of all members of the population is available from which a random sample may be selected.

A random sample of farms on which all relevant animals are tested would be a cluster sample in which the animals on each farm constitute a cluster. While such a sample of farms may be easy to select and convenient for field work, care needs to be exercised in drawing inferences about the disease situation in the animal population, as distinct from the population of farms, from the results. The information obtained will be less accurate than if a sample of the same size, drawn systematically or at random from a list of individual animals, is surveyed. This is because disease statistics, particularly for infectious diseases, are in general more variable between herds than within herds.

Cluster sampling may be applicable to some production line sampling programs such as Salmonella testing of meat products.

3.5 Multistage Sampling

Multistage sampling is a term applied to the selection of a sample in two or more stages. An example would be to randomly select a sample of dairy herds in a State and then to select a random sample of heifers from each of these herds for testing.

4. TERMS AND CONCEPTS

4.1 Prevalence and Incidence

The terms prevalence and incidence are often confused. Frequently the term 'incidence' is used when clearly what is meant is point prevalence.

The point prevalence of a disease is the proportion of a specified population with the disease at a particular point in time. Thus, in most field surveys, it is the point prevalence that is being determined.

On the other hand, the incidence of a disease is the number of new cases of the disease that occurred during a specified period of time. It is a measure of the rapidity with which a disease is occurring or spreading and is usually expressed as a fraction or percentage of the average number of animals in the population during the time period.

Much of the confusion between prevalence and incidence arises because of a third term, period prevalence. Period prevalence is the total number of cases both new and old that have existed at some time during a specified period of time. It is thus a combination of the point prevalence at the beginning of the period and the incidence of new cases during the period. The period prevalence is of limited usefulness in characterising a disease and is most frequently used in describing the pattern of mild, self-limiting diseases.

4.2 Ratios, Proportions and Rates

'Epidemiology is the science of denominators and, as such, is the rational counter-balance of clinical training, which tends to be preoccupied with numerators. The introduction of the denominator was about as important in medical thinking as the invention of the wheel, and equally revolutionary. The denominator is our foundation for rates, and hence for our sense of proportion and priorities. Denominators are dull but indispensible whenever and wherever we try to draw conclusions about distributions, differences, and dividends – fiscally, socially or medically.' (Stewart, 1970).

In epidemiology, as in most scientific disciplines, there is often a certain lack of precision leading to ambiguity in the terminology used. One reason for this may be that workers in one field borrow terms from another field in which they are not specialists. Another reason is semantics; a word may have more than one meaning in common usage or people may use two words as synonyms when, in fact, their meanings are distinct.

The use of the word rate in epidemiology suffers from these disadvantages; it is borrowed from physics and misinterpreted; it has more than one meaning in the English language; and the most common error – it is used inter-changeably with the term proportion because both are incorrectly assumed to be synonyms for ratio.

Ratio is the expression of the relationship between a numerator and denominator where the two are separate and distinct quantities; neither is included in the other. Usually the numerator and denominator are measured in the same units, although this is not essential. For example,

sex ratio = number of males : number of females

$$\text{foetal death ratio} = \frac{\text{number of foetal deaths}}{\text{number of live births}}$$

An index is a comparative measure of two characteristics frequently expressed as a ratio. For example, weight-height index is a ratio used as a measure of obesity.

Proportion, like a ratio, is a relative frequency, but in this case the numerator is included in the denominator, i.e. the proportion some group is of the whole population. For example,

$$\text{proportion of males} = \frac{\text{number of males}}{\text{number of males} + \text{number of females}}$$

$$\text{proportion of foetal deaths} = \frac{\text{number of foetal deaths}}{\text{number of conceptions}}$$

These relative frequencies can be used as an estimate of the probability of an event occurring.

Rate is an expression for the change in one quantity per unit of another quantity, which is frequently time. However, in many biological processes it is not the absolute change per unit time that is of interest, but the relative change per unit of time and per unit of the organism or population undergoing the change - the relative rate. However, since this kind of rate is the most commonly used, the word 'relative' is frequently omitted unless ambiguity would arise from its omission.

Incidence is an example of a rate. Since incidence is the number of new cases in a time period, it is clearly a rate. It does not refer to the population size. We should more properly use relative incidence to refer to the number of new cases per 1000 animals per year. However, by convention we also refer to this as 'incidence', and let the way in which the result is expressed indicate whether we are dealing with a relative rate or not.

Care needs to be taken in selecting the denominator for a relative rate. It is usual to use the mid-period population size rather than the size at the start or end of the period. In a stable situation this may not matter, but, in an epidemic there can be considerable differences between the rates calculated with different denominators. In such cases it may be better to calculate the incidence for shorter periods. Incidence over a short period is generally referred to as the attack rate.

By contrast, prevalence, which is the number of cases at a given time expressed as a proportion of the total population at that time, is clearly a proportion and not a rate. The term 'prevalence rate' is erroneous.

4.3 Sensitivity and Specificity

Most field surveys will involve the use of diagnostic tests (generally serological) for the disease in question. In deciding on the test procedures to be used in a large scale survey, some compromise in regard to accuracy for the sake of simplicity, cost, and acceptability of the test procedure will frequently have to be made. However, there is little point in conducting a large scale survey involving considerable effort in obtaining data if it is not known what the test results mean in terms of the true disease situation. The inferences that can be drawn about the disease situation depend on the characteristics of the diagnostic test being used. These characteristics can be defined in terms of repeatability, accuracy, sensitivity and specificity.

Repeatability (sometimes also referred to as precision) is the ability of a test to give consistent results in repeated tests. Repeatability is the converse of variability.

Accuracy is the ability of a test to give a true measure of the item being tested. Accuracy and repeatability are not the same thing. A test can be repeatable without being accurate, but it cannot be accurate without being repeatable. Accuracy has two components, sensitivity and specificity.

Sensitivity is the ability of a test to give a positive result when the animal is diseased. It is measured as the proportion of animals with the disease that give a positive test result.

Specificity is the ability of a test to give a negative result when the animal is not diseased. It is measured as the proportion of disease-free animals that give a negative test result.

Lack of sensitivity leads to false negative results and lack of specificity leads to false positive results.

It is difficult, however, to accurately assess the sensitivity and specificity of a serological test. In general, the sensitivity of a serological test can be estimated by determining the proportion of positive results in those animals that are positive on microbial culture. However, the specificity of a test cannot generally be assessed by determining the proportion of negative results among animals which are negative on microbial culture. This is not only because some animals negative on culture are in fact infected, but, more importantly, because of the very large number of animals it would be necessary to autopsy and take multiple cultures from to accurately determine anything but a very high proportion of false positive results.

Often, in an attempt to assess sensitivity and specificity, the results of one test are compared with those of another test. This does not establish sensitivity and specificity, but only <u>relative</u> sensitivity and <u>relative</u> specificity. Such comparisons are only of value if the sensitivity and specificity of the standard test are known and they approach 100%.

For any given test, sensitivity and specificity are usually inversely related. Thus if the interpretation of test results is altered to increase the sensitivity (to reduce the number of false negative results), then the specificity decreases (and more false positives will result).

The relationship between sensitivity and specificity can be illustrated by the following table:

		health status		
		diseased	not diseased	total
test result	positive	a	b	a+b
	negative	c	d	c+d
	total	a+c	b+d	a+b+c+d = N

a: diseased animals detected by the test (true positives)
b: non-diseased animals positive to the test (false positives)
c: diseased animals not detected by the test (false negatives)
d: non-diseased animals negative to the test (true negatives)
N: total population

sensitivity = $a/(a+c)$
specificity = $d/(b+d)$

prevalence = $(a+c)/N$
apparent prevalence = $(a+b)/N$

predictive value of a positive result = $a/(a+b)$
predictive value of a negative result = $d/(c+d)$

From this it can be seen that the apparent prevalence observed in a survey will differ from the true prevalence in accordance with the proportion of false positive and false negative results that occur.

An additional term has also been introduced with this table. The predictive value of a positive result (sometimes also referred to as the validity of the test) is the proportion of the test-positive animals (reactors) that really have the disease. The predictive value is of limited use in comparing tests or interpreting the results of a field survey since it is greatly affected by control measures directed at the disease in question. Although the sensitivity and specificity of a test may be constant during a disease control program, while the true prevalence of the disease is being reduced it is likely that the incidence of false positive animals in the population will be unchanged by the control program. It follows that the observed prevalence of the disease will be comprised of an increasing proportion of false positive reactors as the program progresses. Under these conditions the predictive value of a positive result declines markedly.

4.4 Screening, Surveying, Monitoring and Surveillance

The following definitions are included for the sake of completing this glossary and in an attempt to introduce some precision in the use of these terms.

Screening refers to the testing of a wide cross-section of a population to detect new cases of a disease. The test used usually does not aim at establishing a definitive diagnosis, and so the specificity of the test is not of great importance, and can be compromised for the sake of achieving high sensitivity. The test is usually cheap and easily performed.

The term 'screening test' is also used to refer to a test applied to all samples as a means of cost saving; the negatives are treated as disease free, while the positives undergo a more expensive and/or complex test as a definitive criterion.

Surveying refers to the testing of a selected sample of a population to determine the prevalence of disease within that population.

Surveillance is the continuous observation of a population aimed at early case-finding for disease control purposes. It generally involves the testing of some section of the population that is readily accessible, e.g. abattoir surveillance or milk ring testing for brucellosis. While the results of disease surveillance may indicate trends in the disease status of a population, they cannot generally be taken as absolute measures of the incidence or prevalence of the disease in the population as a whole unless the surveillance is based on random sampling.

Monitoring is an on-going testing program aimed at the early detection of changes in the prevalence of disease which in turn might indicate a change in the incidence of that disease. Monitoring generally carries the implication of not being accompanied by control measures when positives are detected, but may stimulate control action where monitoring reveals an increasing prevalence.

Table 1 : Detecting the Presence of a Disease

Tables 1(a), 1(b) and 1(c) give 90%, 95% and 99% confidence levels respectively.

These tables may be used in two ways:

(i) To determine the size of the sample (n) that must be tested to have a given confidence (d) of determining whether a disease is present at a minimum prevalence (d/N) in a population of (N) animals.

- Use the table for the confidence level desired.
- Choose the column corresponding to the percentage of infected animals to be tested against.
- Run down the column to the row that is equal to or greater than the population size to find the required sample size.

(ii) Having tested a random proportion (n/N) of animals in a population of size (N) and found no positives, the upper limit to the number of positives (d) that could be present can be determined.

- Use the table for the confidence level desired.
- Choose the column corresponding to the proportion of the population sampled.
- Read off the upper limit to the possible number of positives from the row corresponding to the population size.

Other values

For values not in the table, either interpolation or the following approximation can be used.

If N is the population size,
d is the number of positives in the population,
n is the number sampled
and α is the desired confidence level (that is, the probability of finding at least one positive in the sample)

Then #Sampled

for (i) $n \simeq \left(1 - (1-\alpha)^{1/d}\right)\left(N - \frac{d}{2}\right) + 1$

and (ii) $d \simeq \left(1 - (1-\alpha)^{1/n}\right)\left(N - \frac{n}{2}\right) + 1$

Although not needed here, it is useful to note that

$$\alpha \simeq 1 - \left(1 - \frac{d}{N - (n-1)/2}\right)^{n}$$

TABLE 1(a)

(i) SAMPLE SIZE REQUIRED FOR DETECTING DISEASE

(ii) CONFIDENCE LIMITS FOR NUMBER OF POSITIVES

90%

	(i) percentage of diseased animals in population (d/N) OR (ii) percentage sampled and found clean (n/N)											
population size (N)	50%	40%	30%	25%	20%	15%	10%	5%	2%	1%	0.5%	0.1%
10	3	4	5	6	7	8	9	10	10	10	10	10
20	4	5	6	7	9	11	14	18	20	20	20	20
30	4	5	6	8	9	12	16	24	30	30	30	30
40	4	5	6	8	10	12	17	27	38	40	40	40
50	4	5	7	8	10	13	18	30	45	50	50	50
60	4	5	7	8	10	13	19	32	52	59	60	60
70	4	5	7	8	10	13	19	34	57	68	70	70
80	4	5	7	8	10	13	20	35	61	76	80	80
90	4	5	7	8	10	14	20	36	65	84	90	90
100	4	5	7	8	10	14	20	37	69	91	100	100
120	4	5	7	8	10	14	20	38	74	103	118	120
140	4	5	7	8	10	14	21	39	78	113	135	140
160	4	5	7	8	10	14	21	40	62	122	152	160
180	4	5	7	8	11	14	21	40	85	130	167	180
200	4	5	7	8	11	14	21	41	87	137	180	200
250	4	5	7	8	11	14	21	42	92	151	211	250
300	4	5	7	8	11	14	22	42	95	161	236	300
350	4	5	7	8	11	14	22	43	98	169	256	350
400	4	5	7	8	11	14	22	43	100	175	274	400
450	4	5	7	8	11	14	22	43	101	180	288	448
500	4	5	7	8	11	14	22	43	102	184	301	496
600	4	5	7	8	11	15	22	44	104	191	321	568
700	4	5	7	8	11	15	22	44	106	196	337	675
800	4	5	7	8	11	15	22	44	107	200	350	756
900	4	5	7	8	11	15	22	44	108	203	360	831
1000	4	5	7	8	11	15	22	44	108	205	369	900
1200	4	5	7	8	11	15	22	45	109	209	382	1024
1400	4	5	7	8	11	15	22	45	110	212	392	1130
1600	4	5	7	8	11	15	22	45	111	214	400	1221
1800	4	5	7	8	11	15	22	45	111	216	406	1299
2000	4	5	7	8	11	15	22	45	111	217	411	1368
3000	4	5	7	8	11	15	22	45	112	221	426	1607
4000	4	5	7	8	11	15	22	45	113	223	434	1750
5000	4	5	7	8	11	15	22	45	113	224	439	1845
6000	4	5	7	8	11	15	22	45	113	225	443	1912
7000	4	5	7	8	11	15	22	45	114	226	445	1962
8000	4	5	7	8	11	15	22	45	114	226	447	2000
9000	4	5	7	9	11	15	22	45	114	227	448	2031
10000	4	5	7	9	11	15	22	45	114	227	449	2056
∞	4	5	7	8	11	15	22	45	114	230	460	2302

The table gives:

(i) the sample size (n) required to be 90% certain of including at least one positive if the disease is present at the specified level

(ii) the upper limit to the number (d) of diseased animals in a population given that the specified proportion were tested and found to be negative.

Examples:

(i) Expected proportion of positives is 2%.
The population size is 480 - use 500.
From the table, a sample of 102 is required to be 90% certain of detecting at least one positive.

(ii) For a population of 1000, a sample of 10% were all found to be negative. From the table, the 90% confidence limit for the number of positives is 22.

TABLE 1(b)

95%

(i) SAMPLE SIZE REQUIRED FOR DETECTING DISEASE

(ii) CONFIDENCE LIMITS FOR NUMBER OF POSITIVES

population size (N)	(i) percentage of diseased animals in population (d/N) OR (ii) percentage sampled and found clean (n/N)											
	50%	40%	30%	25%	20%	15%	10%	5%	2%	1%	0.5%	0.1%
10	4	5	6	7	8	10	10	10	10	10	10	10
20	4	6	7	9	10	12	16	19	20	20	20	20
30	4	6	8	9	11	14	19	26	30	30	30	30
40	5	6	8	10	12	15	21	31	40	40	40	40
50	5	6	8	10	12	16	22	35	48	50	50	50
60	5	6	8	10	12	16	23	38	55	60	60	60
70	5	6	8	10	13	17	24	40	62	70	70	70
80	5	6	8	10	13	17	24	42	68	79	80	80
90	5	6	8	10	13	17	25	43	73	87	90	90
100	5	6	9	10	13	17	25	45	78	96	100	100
120	5	6	9	10	13	18	26	47	86	111	120	120
140	5	6	9	11	13	18	26	48	92	124	139	140
160	5	6	9	11	13	18	27	49	97	136	157	160
180	5	6	9	11	13	18	27	50	101	146	174	180
200	5	6	9	11	13	18	27	51	105	155	190	200
250	5	6	9	11	14	18	27	53	112	175	228	250
300	5	6	9	11	14	18	28	54	117	189	260	300
350	5	6	9	11	14	18	28	54	121	201	287	350
400	5	6	9	11	14	19	28	55	124	211	311	400
450	5	6	9	11	14	19	28	55	127	218	331	450
500	5	6	9	11	14	19	28	56	129	225	349	500
600	5	6	9	11	14	19	28	56	132	235	379	597
700	5	6	9	11	14	19	28	57	134	243	402	691
800	5	6	9	11	14	19	28	57	136	249	421	782
900	5	6	9	11	14	19	28	57	137	254	437	868
1000	5	6	9	11	14	19	29	57	138	258	450	950
1200	5	6	9	11	14	19	29	57	140	264	471	1102
1400	5	6	9	11	14	19	29	58	141	269	487	1236
1600	5	6	9	11	14	19	29	58	142	272	499	1354
1800	5	6	9	11	14	19	29	58	143	275	509	1459
2000	5	6	9	11	14	19	29	58	143	277	517	1553
3000	5	6	9	11	14	19	29	58	145	284	542	1895
4000	5	6	9	11	14	19	29	58	146	288	556	2108
5000	5	6	9	11	14	19	29	59	147	290	564	2253
6000	5	6	9	11	14	19	29	59	147	291	569	2358
7000	5	6	9	11	14	19	29	59	147	292	573	2437
8000	5	6	9	11	14	19	29	59	147	293	576	2498
9000	5	6	9	11	14	19	29	59	148	294	579	2548
10000	5	6	9	11	14	19	29	59	148	294	581	2588
∞	5	6	9	11	14	19	29	59	149	299	598	2995

The table gives:

(i) the sample size (n) required to be 95% certain of including at least one positive if the disease is present at the specified level

(ii) the upper limit to the number (d) of diseased animals in a population given that the specified proportion were tested and found to be negative.

Examples:

(i) Expected proportion of positives is 2%.
The population size is 480 - use 500.
From the table, a sample of 129 is required to be 95% certain of detecting at least one positive.

(ii) For a population of 1000, a sample of 10% were all found to be negative. From the table, the 95% confidence limit for the number of positives is 29.

TABLE 1(c)

99%

(i) SAMPLE SIZE REQUIRED FOR DETECTING DISEASE

(ii) CONFIDENCE LIMITS FOR NUMBER OF POSITIVES

	(i) percentage of diseased animals in population (d/N) OR (ii) percentage sampled and found clean (n/N)											
population size (N)	50%	40%	30%	25%	20%	15%	10%	5%	2%	1%	0.5%	0.1%
10	5	6	7	10	10	10	10	10	10	10	10	10
20	6	8	10	11	13	15	18	20	20	20	20	20
30	6	8	11	13	15	19	23	30	30	30	30	30
40	7	8	11	13	16	21	27	36	40	40	40	40
50	7	9	12	14	17	22	29	42	50	50	50	50
60	7	9	12	14	18	23	31	47	60	60	60	60
70	7	9	12	15	18	24	33	51	68	70	70	70
80	7	9	12	15	19	24	34	54	76	80	80	80
90	7	9	12	15	19	25	35	57	83	90	90	90
100	7	9	13	15	19	25	36	59	90	100	100	100
120	7	9	13	15	19	26	37	63	102	118	120	120
140	7	9	13	16	20	26	38	67	113	135	140	140
160	7	9	13	16	20	26	39	69	122	151	160	160
180	7	9	13	16	20	27	39	71	129	166	179	180
200	7	9	13	16	20	27	40	73	136	180	198	200
250	7	9	13	16	20	27	40	76	150	210	244	250
300	7	9	13	16	20	27	41	78	160	235	286	300
350	7	9	13	16	21	28	42	80	168	256	325	350
400	7	9	13	16	21	28	42	81	174	273	360	400
450	7	9	13	16	21	28	42	82	179	288	392	450
500	7	9	13	16	21	28	42	83	183	300	421	500
600	7	9	13	16	21	28	43	84	190	321	470	600
700	7	9	13	16	21	28	43	85	195	336	512	700
800	7	9	13	16	21	28	43	85	199	349	546	798
900	7	9	13	16	21	28	43	86	202	359	576	895
1000	7	9	13	16	21	28	43	86	204	368	601	990
1200	7	9	13	16	21	28	43	87	208	381	642	1175
1400	7	9	13	16	21	29	44	87	211	391	674	1348
1600	7	9	13	16	21	29	44	88	213	399	699	1510
1800	7	9	13	16	21	29	44	88	215	405	720	1661
2000	7	9	13	16	21	29	44	88	216	410	737	1800
3000	7	9	13	16	21	29	44	89	220	425	792	2353
4000	7	10	13	16	21	29	44	89	222	433	821	2735
5000	7	10	13	16	21	29	44	89	223	438	840	3009
6000	7	10	13	16	21	29	44	90	224	442	852	3214
7000	7	10	13	16	21	29	44	90	225	444	861	3373
8000	7	10	13	16	21	29	44	90	225	446	868	3500
9000	7	10	13	16	21	29	44	90	226	447	874	3604
10000	7	10	13	16	21	29	44	90	226	448	878	3689
∞	7	10	13	16	21	29	44	90	228	459	919	4603

The table gives:

(i) the sample size (n) required to be 99% certain of including at least one positive if the disease is present at the specified level

(ii) the upper limit to the number (d) of diseased animals in a population given that the specified proportion were tested and found to be negative.

Examples:

(i) Expected proportion of positives is 2%.
The population size is 480 - use 500.
From the table, a sample of 183 is required to be 99% certain of detecting at least one positive.

(ii) For a population of 1000, a sample of 10% were all found to be negative. From the table, the 99% confidence limit for the number of positives is 43.

Table 2: Chances of Detecting Positives With Various Intensities of Monitoring

These tables show the effect of different levels of monitoring on the chance of detecting infection that is present at a low level.

Tabulated is the probability of finding at least one positive, given that we are sampling at one of 5 levels (20%, 30%, 40%, 50%, 60%) for populations with a size from 10 to 100, and containing 1 to 8 positives. Under these conditions, with population size greater than 100, the probabilities quickly reach their asymptotic limit for large populations, and this is also given.

Other values can easily be calculated from the approximation mentioned in the notes to Table 1.

Intensities of

nitoring on the
.evel.

at least one
vels (20%, 30%,
10 to 100, and
with population
y reach their
o given.

he approximation

TABLE 2: THE CHANCES OF DETECTING A SMALL NUMBER OF POSITIVES

These tables give the probability of detecting at least one positive for different sampling intensites and numbers of positives in the population.

TABLE 2(a): 20% SAMPLING

population size	number sampled	number of positives in the population 1	2	3	4	5	6	7	8
10	2	0.200	0.378	0.533	0.667	0.778	0.867	0.933	0.978
20	4	0.200	0.368	0.509	0.624	0.718	0.793	0.852	0.898
30	6	0.200	0.366	0.501	0.612	0.702	0.773	0.830	0.874
40	8	0.200	0.364	0.496	0.607	0.694	0.764	0.815	0.863
50	10	0.200	0.363	0.496	0.603	0.689	0.758	0.813	0.857
60	12	0.200	0.363	0.495	0.601	0.686	0.755	0.809	0.853
70	14	0.200	0.362	0.494	0.599	0.684	0.752	0.807	0.850
80	16	0.200	0.362	0.493	0.598	0.683	0.751	0.804	0.847
90	18	0.200	0.362	0.492	0.597	0.682	0.749	0.803	0.846
100	20	0.200	0.362	0.492	0.597	0.681	0.748	0.802	0.844
∞	∞	0.200	0.360	0.488	0.590	0.672	0.738	0.790	0.832

TABLE 2(b): 30% SAMPLING

population size	number sampled	number of positives in the population 1	2	3	4	5	6	7	8
10	3	0.300	0.533	0.708	0.833	0.917	0.967	0.992	0.994
20	6	0.300	0.521	0.681	0.793	0.871	0.923	0.956	0.976
30	9	0.300	0.517	0.672	0.782	0.857	0.909	0.943	0.965
40	12	0.300	0.515	0.668	0.776	0.851	0.902	0.936	0.960
50	15	0.300	0.514	0.666	0.773	0.847	0.898	0.933	0.956
60	18	0.300	0.514	0.665	0.770	0.844	0.895	0.930	0.954
70	21	0.300	0.513	0.663	0.769	0.842	0.893	0.928	0.952
80	24	0.300	0.513	0.663	0.768	0.841	0.892	0.927	0.951
90	27	0.300	0.512	0.662	0.767	0.840	0.891	0.926	0.950
100	30	0.300	0.512	0.661	0.766	0.839	0.890	0.925	0.949
∞	∞	0.300	0.510	0.657	0.760	0.832	0.882	0.918	0.942

TABLE 2(c): 40% SAMPLING

population size	number sampled	number of positives in the population 1	2	3	4	5	6	7	8
10	4	0.400	0.667	0.833	0.929	0.976	0.994	0.994	0.994
20	8	0.400	0.653	0.807	0.898	0.949	0.976	0.990	0.994
30	12	0.400	0.646	0.795	0.886	0.940	0.969	0.984	0.993
40	16	0.400	0.646	0.795	0.884	0.935	0.965	0.981	0.990
50	20	0.400	0.645	0.793	0.881	0.933	0.963	0.980	0.989
60	24	0.400	0.644	0.791	0.879	0.931	0.961	0.978	0.988
70	28	0.400	0.643	0.790	0.878	0.930	0.960	0.977	0.987
80	32	0.400	0.643	0.789	0.877	0.929	0.959	0.977	0.987
90	36	0.400	0.643	0.789	0.876	0.928	0.959	0.976	0.987
100	40	0.400	0.642	0.788	0.876	0.927	0.958	0.976	0.986
∞	∞	0.400	0.640	0.784	0.870	0.922	0.953	0.972	0.983

TABLE 2(d): 50% SAMPLING

population size	number sampled	number of positives in the population 1	2	3	4	5	6	7	8
10	5	0.500	0.778	0.917	0.976	0.994	0.994	0.994	0.994
20	10	0.500	0.763	0.895	0.957	0.984	0.994	0.994	0.994
30	15	0.500	0.759	0.888	0.950	0.979	0.992	0.994	0.994
40	20	0.500	0.756	0.885	0.947	0.976	0.990	0.994	0.994
50	25	0.500	0.755	0.883	0.945	0.975	0.989	0.994	0.994
60	30	0.500	0.754	0.881	0.944	0.974	0.986	0.994	0.994
70	35	0.500	0.754	0.880	0.943	0.973	0.986	0.994	0.994
80	40	0.500	0.753	0.880	0.942	0.973	0.987	0.994	0.994
90	45	0.500	0.753	0.879	0.942	0.972	0.987	0.994	0.994
100	50	0.500	0.753	0.875	0.941	0.972	0.987	0.994	0.994
∞	∞	0.500	0.750	0.875	0.937	0.969	0.984	0.992	0.996

TABLE 2(e): 60% SAMPLING

population size	number sampled	number of positives in the population 1	2	3	4	5	6	7	8
10	6	0.600	0.867	0.967	0.994	0.994	0.994	0.994	0.994
20	12	0.600	0.853	0.951	0.986	0.994	0.994	0.994	0.994
30	18	0.600	0.848	0.946	0.982	0.994	0.994	0.994	0.994
40	24	0.600	0.846	0.943	0.980	0.993	0.994	0.994	0.994
50	30	0.600	0.845	0.942	0.979	0.993	0.994	0.994	0.994
60	36	0.600	0.844	0.941	0.978	0.992	0.994	0.994	0.994
70	42	0.600	0.843	0.940	0.978	0.992	0.994	0.994	0.994
80	48	0.600	0.843	0.940	0.977	0.992	0.994	0.994	0.994
90	54	0.600	0.843	0.939	0.977	0.991	0.994	0.994	0.994
100	60	0.600	0.842	0.939	0.977	0.991	0.994	0.994	0.994
∞	∞	0.600	0.840	0.936	0.974	0.990	0.996	0.998	0.995

Example:

A 40% sample from a herd of 20 animals would have a 98% chance of including at least one positive if 6 were present in the herd.

Table 3: Probability of Failure to Detect Diseased Animals

The table gives the probability of failure to detect diseased animals from an 'infinite' population with the specified proportion of positives in the population.

prevalence	number of animals in sample tested 5	10	25	50	75	100	200	250	500	1[illegible]
1%	0.951	0.904	0.778	0.605	0.471	0.366	0.134	0.081	0.007	0.[illegible]
2%	0.904	0.817	0.603	0.364	0.220	0.133	0.018	0.006	0.000	
3%	0.859	0.737	0.467	0.218	0.102	0.048	0.002	0.000		
4%	0.815	0.665	0.360	0.130	0.047	0.017	0.000			
5%	0.774	0.599	0.277	0.077	0.021	0.006	0.000			
6%	0.734	0.539	0.213	0.045	0.010	0.002	0.000			
7%	0.696	0.484	0.163	0.027	0.004	0.001	0.000			
8%	0.659	0.434	0.124	0.015	0.002	0.000				
9%	0.624	0.389	0.095	0.009	0.001	0.000				
10%	0.590	0.349	0.072	0.005	0.000					
12%	0.528	0.279	0.041	0.002	0.000					
14%	0.470	0.221	0.023	0.001	0.000					
16%	0.418	0.175	0.013	0.000						
18%	0.371	0.137	0.007	0.000						
20%	0.328	0.107	0.004	0.000						
24%	0.254	0.064	0.001	0.000						
28%	0.193	0.037	0.000							
32%	0.145	0.021	0.000							
36%	0.107	0.012	0.000							
40%	0.078	0.006	0.000							
50%	0.031	0.001	0.000							
60%	0.010	0.000								

While this table is similar to Table 2 in that it shows the effect of sample size on the probability of detecting at least one diseased animal, it approaches the problem differently.

The table assumes a specified number of samples (n) taken from an infinite population with a proportion (θ) of positives. The probability that we fail to detect any positives in the sample is simply:

$$(1 - \theta)^n$$

and it is this function that is tabulated.

Example:

Tests of a series of random samples of 25 animals from a large population in which 10% of animals are positive would fail to detect any positives in 7% of such sample-groups.

Table 4: Sample Size for Estimation of Disease Prevalence

The table gives the approximate sample size required to estimate a prevalence in a large population with the desired fixed width confidence limits.

	level of confidence								
	90% desired accuracy			95% desired accuracy			99% desired accuracy		
expected prevalence	10	5	1	10	5	1	10	5	1
10%	24	97	2435	35	138	3457	60	239	5971
20%	43	173	4329	61	246	6147	106	425	10616
30%	57	227	5682	81	323	8067	139	557	13933
40%	65	260	6494	92	369	9220	159	637	15923
50%	68	271	6764	96	384	9604	166	663	16587
60%	65	260	6494	92	369	9220	159	637	15923
70%	57	227	5682	81	323	8067	139	557	13933
80%	43	173	4329	61	246	6147	106	425	10616
90%	24	97	2435	35	138	3457	60	239	5971

The table assumes a knowledge of the approximate result. If in doubt, either use the 0.5 figure, or use the 0.2 figure, but take additional samples if necessary.

When sampling from a finite population of size N, an adjustment to account for this can be made by calculating the sample size n_∞ above and calculating

$$\frac{1}{n} = \frac{1}{n_\infty} + \frac{1}{N}$$

to give n, the approximate sample size required.

The table is based on the normal approximation to the binomial distribution and Appendix A.2 gives further details.

Another way to estimate the sample size required is to use Table 5 in reverse:

Choose the desired confidence width, and estimate the values of N that correspond to this width at both p and 1-p, where p is the expected proportion. Use the larger of the two. If the population is not large, the same correction can be made as above.

Example:

We want to estimate the prevalence of a disease in a herd of 1127 animals to within 5% at the 95% confidence level. We expect that the prevalence would be about 40%.

From the table we see that 369 animals would need to be sampled from a large population, and we correct this for the 'finite population' by calculating

$$\frac{1}{n} = \frac{1}{369} + \frac{1}{1127} = \frac{1}{278} .$$

Thus testing of 278 animals would be sufficient.

Table 5: Estimation of Disease Prevalence

These graphs give the width of the upper and lower confidence limits for use in estimating disease prevalence from a simple random survey. The confidence limits have been calculated so that the tails are of equal probability. Thus the graphs can be used to find either the 90%, 95% or 99% two-sided limits or the 95%, 97.5% or 99.5% one-sided confidence limits.

- For a sample of size n, in which a proportion p were found to be positive, calculate $q = 1-p$, the proportion negative
- The upper limit to the prevalence estimate is found by reading off the graph for the appropriate p and n, and adding this amount to the estimate p
- The lower limit is found by reading off the graph for the appropriate q and n, and subtracting the amount from the estimate p.

This table has been drawn up for use in estimating prevalence in a very large population. For a small population, the width of the confidence intervals should be multiplied by a finite correction factor $\sqrt{1-f}$, where f is the proportion of the population sampled:

Proportion sampled	(f)	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
Correction factor	($\sqrt{1-f}$)	1	.95	.90	.84	.78	.71	.64	.55	.45	.32	.0

er confidence
simple random
at the tails
find either
.5% or 99.5%

were found to
egative

is found by
n, and adding

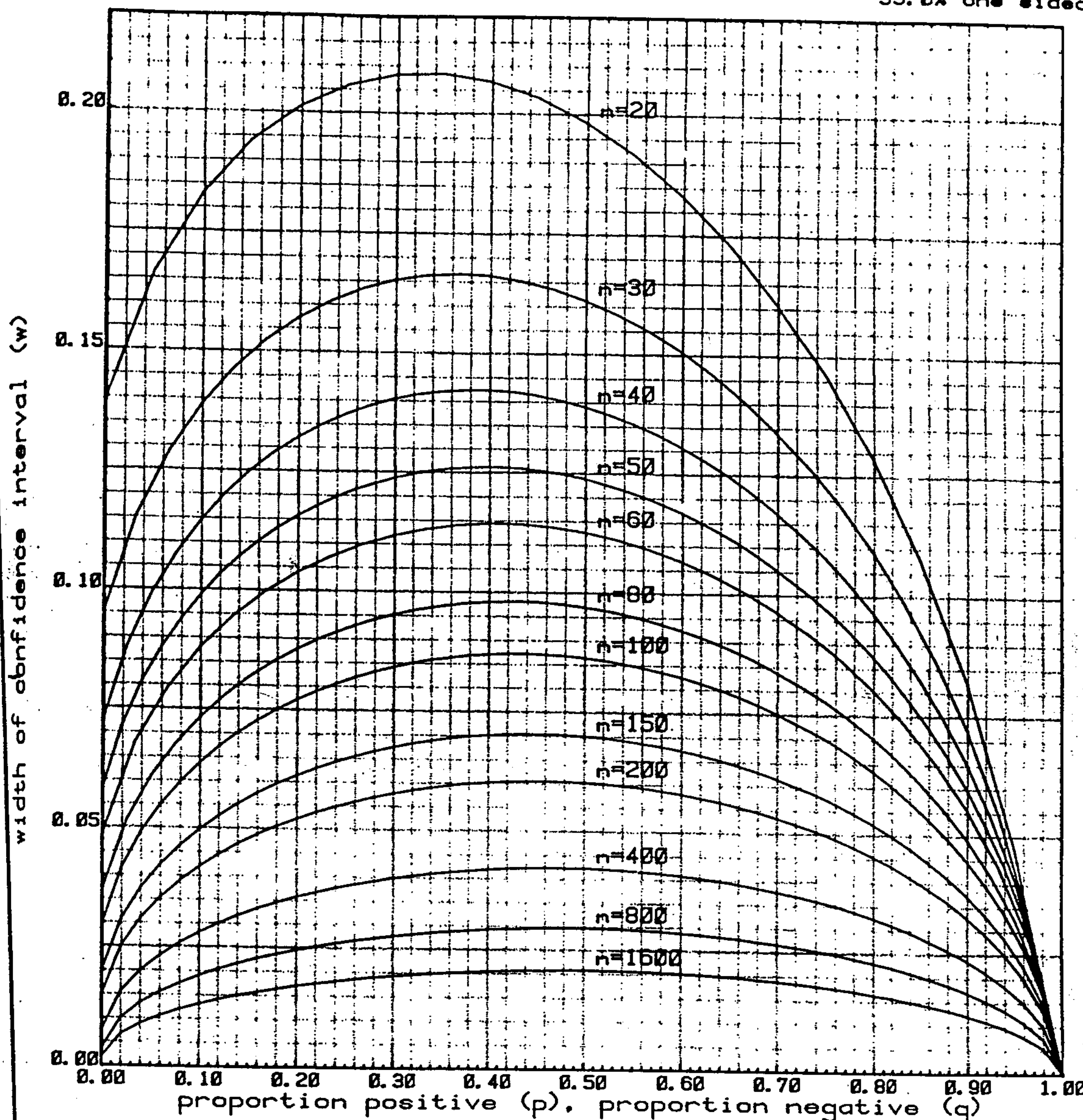
graph for the
mount from the

ting prevalence
he width of the
nite correction
on sampled:

.7	.8	.9	1.0
.55	.45	.32	.0

TABLE 5(a)
BINOMIAL CONFIDENCE LIMITS

90% two sided
95.0% one sided



The table gives the width (w) of two sided 90% confidence limits for a prevalence given an observed proportion (p) of positives with different sample sizes (n).

Example:

In a sample of 60 animals, 12 were found to be positive

Calculate p = 12/60 = .2 and q = 1 - p = .8

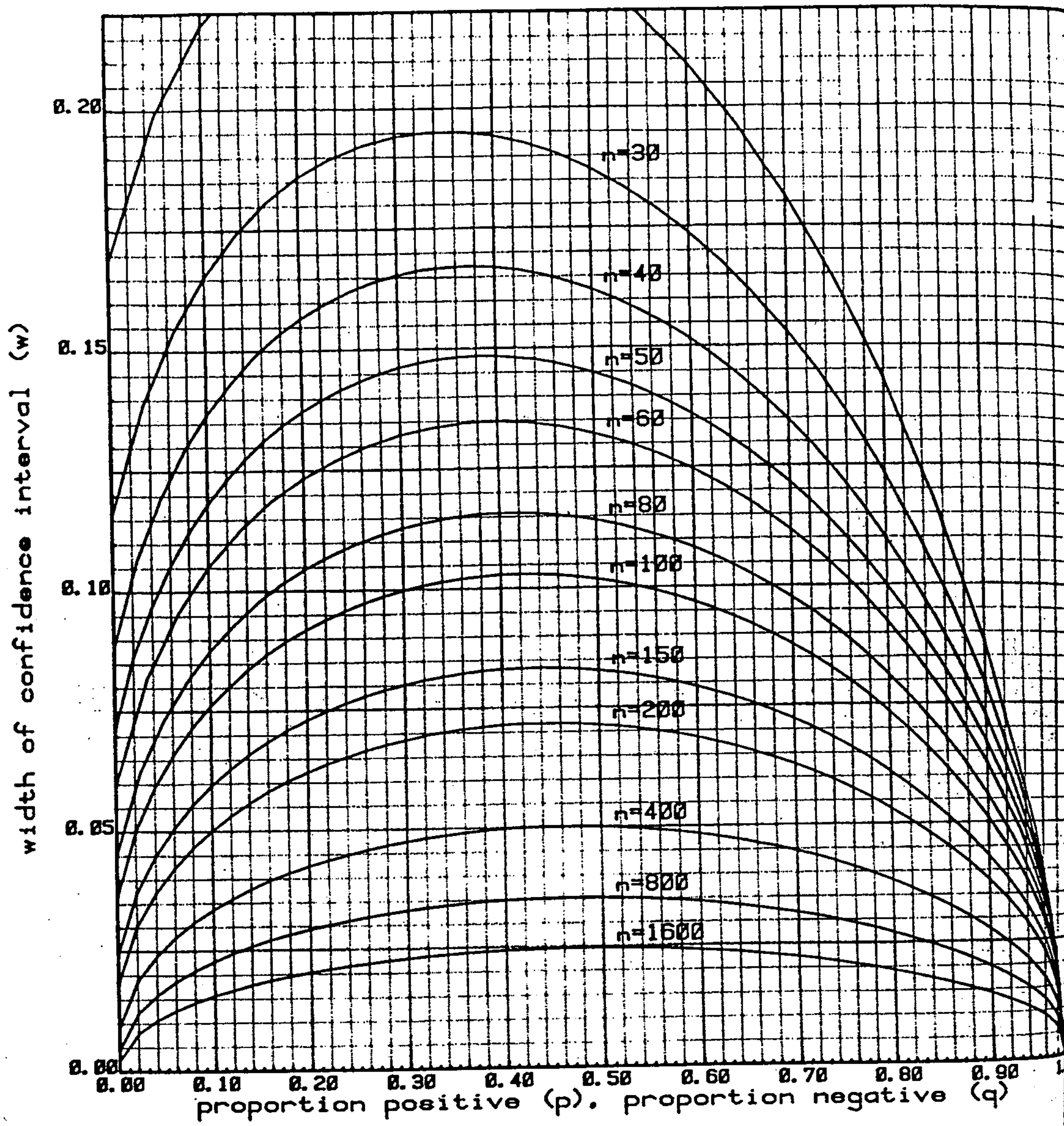
Read off
upper width of 0.104 at p
and lower width of 0.080 at q

.104 + .2 = .304
.2 − .08 = .120

90% confidence limits for the prevalence are (0.120 - 0.304)

TABLE 5(b)
BINOMIAL CONFIDENCE LIMITS

95% two sided
97.5% one sided



The table gives the width (w) of two sided 95% confidence limits for a prevalence given an observed proportion (p) of positives with different sample sizes (n).

Example:

In a sample of 60 animals, 12 were found to be positive

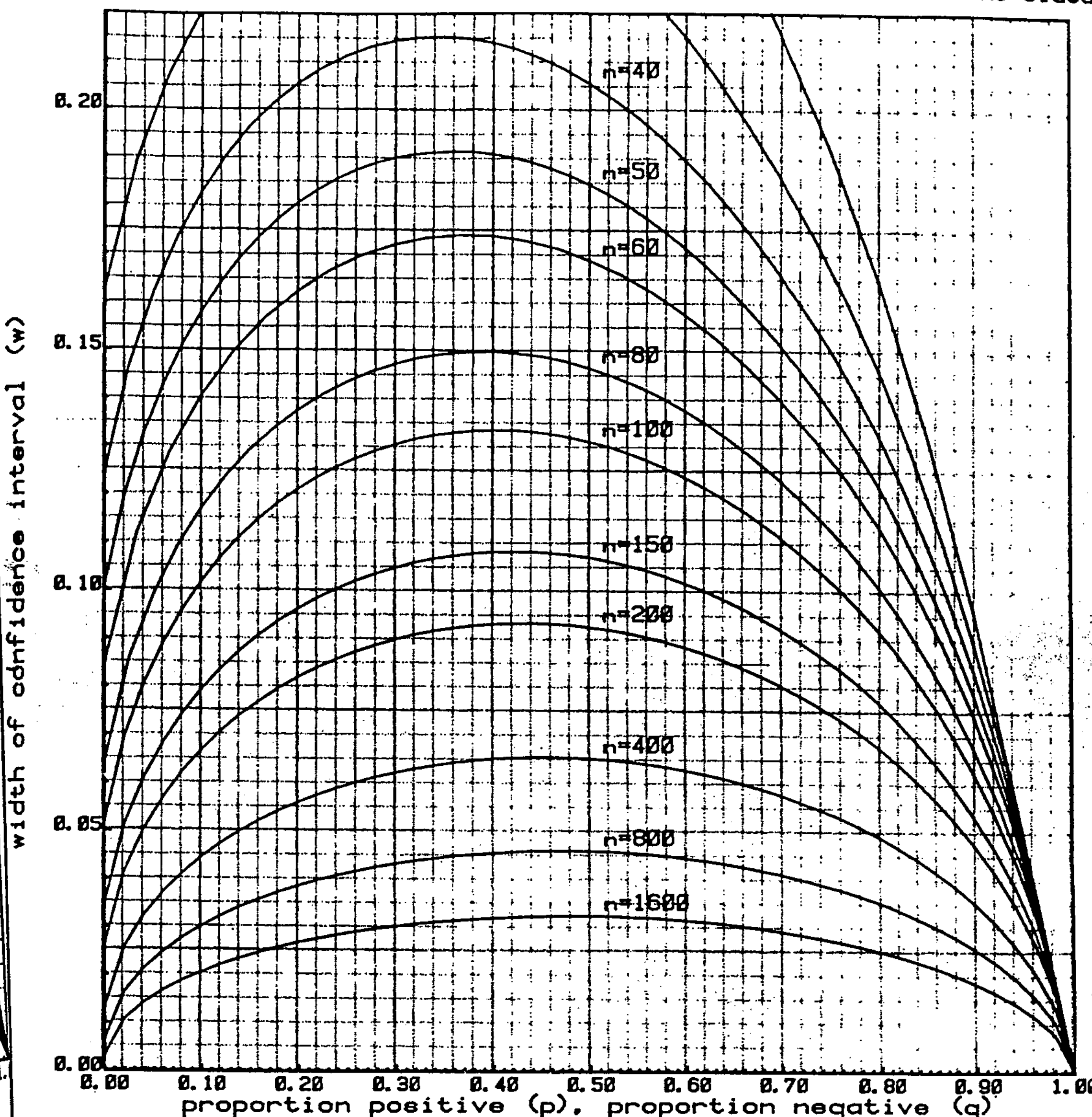
Calculate p = 12/60 = .2 and q = 1 - p = .8

Read off
upper width of 0.123 at p .2+.123=.323
and lower width of 0.092 at q .2-.092=.108

95% confidence limits for the prevalence are (0.108 - 0.323)

99% two sided
99.5% one sided

TABLE 5(c)
BINOMIAL CONFIDENCE LIMITS



The table gives the width (w) of two sided 99% confidence limits for a prevalence given an observed proportion (p) of positives with different sample sizes (n).

Example:

In a sample of 60 animals, 12 were found to be positive

Calculate p = 12/60 = .2 and q = 1 - p = .8

Read off
upper width of 0.162 at p
and lower width of 0.113 at q

99% confidence limits for the prevalence are (0.087 - 0.362)

Table 6: Random Numbers

The table consists of sets of two-digit random numbers. It can be used in selecting a random sample from any population of known size.

To select a sample of a required size, start at any point in the table and move systematically down columns and along rows until a sample of the required size is obtained. Where a number is duplicated in the group selected it should be ignored and the next random number in the sequence used. Similarly if a number falls outside the range of the population size it should be skipped over and next random number used. This does not affect the randomness of the sample.

Examples:

i) To select a random sample of 10 from a herd of 40

- Number the animals 1 to 40
- Start at any point in the random number table
- Work down the column writing down each number that is from 1 to 40 but rejecting repetitions until you have ten numbers.

To avoid 'wasting' random numbers, subtract 40 from any numbers over 40 and use the result if it is 40 or less. But if the result is again over 40, do not repeat this process. The numbers would then only go from 1 to 19 and there would be no chance of selecting 20 to 40.

In the sequence of random numbers below, the ten animals chosen are underlined. Note that both 6 and 13 were duplicated in the list.

random number:	98	05	89	80	23	64	53	60	75	84	06	88	06	11	13	16	81
subtract 40:	58		49	40		24	13	20	35	44		48					

The sample would comprise animals 5, 6, 11, 13, 16, 20, 23, 24, 35 and 40.

ii) To sample a tenth of the sheep on a property where these are in three mobs of 357, 117 and 310 respectively:

- Using the method of (i), random samples of 36, 12 and 31 respectively may be selected from the three mobs to give a stratified sample
- Alternatively use pairs of two-digit random numbers to create a number between 0 and 999 by ignoring the fourth digit.

For random numbers:

1-357 - first paddock, sheep 1-357
401-517 - second paddock, sheep 1-117
601-910 - third paddock, sheep 1-310

Ignore any other numbers.

iii) To sample an eighth of the items from a continuous production process:

- Start at any point in the random number table
- If the last digit in the two-digit random number is 0 or 9 ignore it, and use the next random number
- Sample the item if the last digit is 2.

TABLE 6: RANDOM NUMBERS

A1

85 36	06 90	48 32	40 29	16 68	44 75	18 65	69 79	77 58	17 32	59 57	30 35
52 94	38 68	59 50	59 09	57 13	87 49	65 52	46 08	77 43	76 78	46 38	55 80
29 50	05 63	70 53	74 49	10 31	82 40	83 34	06 38	05 25	93 31	99 85	68 64
83 88	19 16	51 38	82 91	95 07	20 09	74 20	54 49	35 58	47 09	97 04	57 77
90 51	92 16	14 32	66 52	88 20	05 55	25 92	95 10	51 33	54 51	11 09	38 03
60 20	17 24	48 65	28 20	74 60	80 24	75 73	00 74	29 03	62 01	09 40	40 67
63 09	19 86	24 94	98 53	03 87	11 00	80 03	53 73	24 74	03 33	93 08	86 02
64 82	30 48	54 07	31 60	30 54	08 56	06 45	74 97	60 42	10 63	52 77	66 94
98 64	13 47	64 96	03 62	55 23	55 46	08 03	02 55	06 66	38 65	07 07	74 41
85 40	00 09	63 57	53 09	82 73	18 90	49 73	05 04	55 71	65 85	32 97	99 86
60 36	57 53	96 30	10 81	61 07	89 82	59 92	24 18	03 77	61 08	83 64	65 76
14 98	03 81	15 75	44 63	14 15	01 46	76 25	50 81	58 66	19 58	11 06	59 35
75 05	23 52	20 97	25 54	79 97	96 68	03 39	53 54	66 03	66 16	39 84	54 95
11 06	18 55	46 40	62 06	89 33	74 63	57 54	43 18	88 13	10 21	99 82	86 41
92 15	36 49	72 99	74 93	30 37	28 11	34 60	51 45	97 64	32 39	37 64	45 06
83 16	21 65	22 24	13 24	89 85	15 69	24 43	18 02	11 16	66 36	30 94	40 14
11 12	29 72	36 01	25 20	26 33	43 66	42 16	68 77	25 14	08 18	05 02	73 04
08 13	37 10	67 19	15 69	67 82	47 46	87 34	65 56	62 57	54 38	16 16	32 92
57 02	64 42	64 53	99 48	64 73	31 53	75 63	47 48	10 51	56 31	65 03	83 98
01 08	76 51	25 14	05 89	83 12	31 38	48 20	59 70	85 96	24 92	65 43	10 33
00 45	59 03	80 06	85 50	56 09	82 95	95 01	74 18	73 39	97 25	57 09	35 34
64 37	68 72	66 29	49 36	86 45	43 70	27 10	85 40	17 28	89 94	39 89	29 31
71 43	66 59	30 99	75 82	57 35	15 97	38 97	14 67	12 52	46 20	39 85	05 00
70 64	57 58	26 45	10 88	48 51	43 82	12 90	39 20	14 70	99 61	04 50	45 36
02 80	30 15	98 42	23 92	37 47	34 39	32 91	27 23	94 43	73 36	13 07	99 62
31 09	01 26	91 76	12 99	94 94	31 29	73 15	20 10	87 44	28 76	12 35	50 18
02 78	94 89	22 36	93 56	19 44	86 46	19 82	63 23	43 44	63 51	94 26	40 96
40 62	86 74	85 52	05 60	62 42	64 15	60 75	03 50	52 48	12 21	69 20	98 20
85 35	20 52	24 46	40 60	43 29	93 77	33 20	76 47	34 10	25 30	83 89	69 12
67 03	13 47	72 02	21 11	77 07	11 44	25 39	68 44	04 32	77 49	27 20	63 58
05 99	95 46	38 94	89 24	84 24	23 88	82 77	98 06	83 63	65 75	40 76	60 11
59 34	18 06	02 60	34 79	11 27	76 73	98 93	05 88	67 04	49 68	88 35	69 01
59 18	82 27	66 87	00 99	98 17	40 04	72 78	89 06	03 91	43 49	91 08	40 78
09 05	34 87	91 83	94 74	79 04	50 78	17 54	80 11	90 34	87 47	24 55	98 51
45 64	87 99	58 57	64 74	12 90	30 26	43 09	23 13	20 32	79 88	38 30	08 41
39 97	13 69	95 05	46 14	91 70	94 68	02 75	64 16	39 74	25 30	03 89	91 48
90 21	46 51	12 87	00 69	86 77	57 86	15 21	53 81	25 46	38 96	63 04	28 60
18 79	67 55	52 78	00 91	46 71	04 43	57 78	60 97	45 72	00 24	67 43	08 92
72 70	25 83	26 49	88 66	59 78	16 31	20 30	75 52	08 23	39 74	06 53	20 19
48 41	24 67	96 31	97 15	56 93	87 13	28 90	84 16	48 33	92 34	54 16	82 21
14 65	97 52	86 40	48 95	07 09	57 82	02 30	40 26	84 21	12 52	79 08	64 78
82 57	02 78	77 86	71 36	04 16	44 22	01 78	45 67	07 62	99 16	24 94	39 65
74 95	31 06	38 21	20 72	43 76	85 53	41 26	04 33	33 00	80 51	10 17	25 43
92 89	64 97	55 42	84 38	86 24	87 31	54 70	60 24	06 25	13 79	59 87	79 60
61 14	57 05	58 48	65 67	91 06	46 37	63 04	71 81	42 94	80 09	97 89	44 23
93 65	54 34	80 93	80 29	63 29	18 90	79 59	83 75	30 64	94 52	07 06	62 20
73 82	50 10	44 44	84 55	32 46	22 90	26 73	77 16	73 33	62 48	54 12	85 13
35 07	75 94	38 17	15 51	63 97	39 08	02 98	08 19	93 07	43 65	70 99	08 41
49 87	65 35	48 06	85 63	41 94	47 17	64 79	60 92	15 33	84 96	87 03	68 17
55 07	31 64	23 29	33 82	36 45	73 89	65 89	05 63	44 16	00 59	54 27	11 29
33 25	28 65	12 62	09 05	98 62	31 44	11 20	86 53	76 98	32 76	44 58	23 21
03 79	76 72	15 29	28 16	07 33	57 79	56 75	49 87	23 87	70 63	55 90	37 18
86 04	88 56	62 01	25 93	08 98	35 74	85 83	74 97	17 68	27 92	79 06	45 71
48 96	04 43	04 12	01 09	50 52	41 69	75 60	45 17	99 54	83 87	97 56	21 04
04 34	80 62	92 44	85 75	65 16	49 04	92 96	34 92	48 47	98 93	98 30	45 43
94 75	65 44	60 98	32 68	92 26	76 97	84 95	29 37	42 71	93 64	12 22	83 16
26 91	54 07	56 43	63 29	30 36	30 42	24 92	01 57	16 31	59 83	85 58	09 09
78 10	53 05	29 47	36 21	11 85	07 30	71 60	65 47	35 83	96 09	94 08	98 88
24 69	40 70	44 21	83 63	97 38	86 25	77 05	65 28	47 80	47 68	82 71	00 14
92 61	23 07	25 80	43 65	98 85	28 02	68 76	43 63	10 37	62 52	01 13	30 55
59 92	94 17	26 40	80 92	41 80	29 24	97 63	52 18	06 77	91 33	11 32	29 59
97 72	51 51	88 39	27 84	53 20	31 23	03 84	08 22	08 41	91 68	33 13	85 93
25 80	11 89	10 49	29 31	28 63	91 53	98 19	36 29	63 54	89 94	38 71	00 44
03 23	54 29	68 73	94 42	35 50	62 39	56 98	65 45	47 65	48 38	48 08	67 30
37 09	68 82	34 32	60 66	41 26	62 23	34 15	47 35	87 86	46 10	33 53	05 41
65 44	80 16	35 00	34 97	46 75	76 81	39 09	20 94	92 45	87 56	32 27	36 74
95 33	13 86	46 62	09 81	44 09	74 57	57 36	25 48	16 67	15 71	10 36	48 78
31 51	59 38	85 88	07 87	79 89	98 95	45 76	24 53	77 85	39 15	63 56	32 53
74 28	65 63	80 32	40 46	36 18	58 97	51 12	70 56	61 01	64 23	85 51	68 97
94 68	05 36	54 16	48 14	60 78	88 10	29 23	70 60	93 35	06 98	23 90	20 79

10 66	53 13	45 41	18 77	97 10	70 12	36 08	24 76	10 37	01 67	59 45	46 87
74 04	55 13	48 92	56 99	70 39	18 11	76 99	87 46	15 94	22 43	56 15	24 91
18 02	78 59	22 12	77 31	77 08	25 01	96 57	75 26	27 00	31 32	49 31	10 94
97 19	10 79	78 35	10 32	23 35	53 32	69 16	44 21	15 44	04 06	35 81	59 07
50 91	92 01	00 21	74 20	78 13	14 97	80 66	55 71	97 83	98 14	50 00	00 13
73 70	12 12	29 20	01 66	84 45	45 61	13 04	27 96	92 93	06 66	44 03	51 88
86 72	15 57	41 82	13 15	16 39	02 21	67 75	64 21	71 58	86 68	36 32	14 50
10 79	16 94	14 14	27 63	71 03	18 49	48 69	87 61	16 73	72 73	89 89	86 04
81 21	07 60	01 74	74 68	47 94	29 73	90 36	62 86	89 94	43 28	30 13	62 51
73 79	08 78	11 06	44 77	92 76	21 99	07 80	01 36	23 50	33 41	28 69	22 89

50 27	83 26	35 73	57 81	15 55	50 22	97 13	11 74	23 82	82 39	56 47	54 00
42 70	87 50	67 04	48 80	22 91	20 93	20 47	69 19	77 14	52 11	89 12	77 56
13 99	47 30	23 73	90 64	60 34	45 90	47 17	55 50	75 44	07 37	67 86	80 13
72 92	83 11	07 60	99 39	66 48	90 08	08 17	47 52	42 74	93 83	17 39	39 39
60 13	88 18	10 76	55 23	10 54	59 66	69 23	86 98	38 57	53 89	97 90	35 22
23 53	07 75	91 48	74 34	96 90	27 80	69 36	07 11	78 36	09 50	13 74	24 47
09 53	16 47	16 88	30 14	47 33	53 06	05 51	20 73	03 00	30 83	06 00	06 83
61 99	15 33	16 79	83 80	14 20	51 28	41 68	41 48	60 39	44 20	25 93	28 58
19 75	97 23	31 43	37 39	61 17	81 21	28 23	57 07	58 59	33 76	16 58	13 66
42 89	92 39	09 28	49 27	54 04	73 68	94 79	09 79	22 39	65 76	82 60	35 68

41 42	90 11	75 51	54 18	28 60	36 21	59 43	51 86	87 51	12 90	62 19	60 63
66 11	53 82	38 11	32 67	95 91	69 51	43 23	01 48	43 80	84 02	53 10	53 45
11 67	95 97	89 73	58 11	22 74	64 09	90 00	28 10	20 09	24 06	64 69	19 32
65 24	18 74	82 92	70 60	68 34	34 13	10 00	27 65	27 32	36 93	98 66	62 77
77 30	65 22	71 77	74 20	48 84	26 85	66 23	02 91	98 65	63 22	87 02	02 55
61 14	42 43	08 79	45 38	01 25	88 63	38 94	27 04	38 43	10 77	58 44	42 72
47 52	04 05	88 12	12 39	22 96	56 02	02 38	73 33	32 76	89 73	21 02	79 32
98 23	15 58	16 11	41 17	30 48	01 07	93 31	09 41	66 54	98 36	70 83	13 40
29 28	79 94	58 26	38 86	55 49	87 28	34 99	57 92	24 73	29 18	90 47	40 17
66 57	28 41	17 21	98 43	35 99	18 54	33 84	01 34	93 29	55 31	48 22	35 13

37 26	02 34	03 45	10 27	27 99	92 49	23 12	06 08	42 06	15 35	84 37	93 60
33 20	29 33	60 91	70 40	06 97	49 91	02 35	47 15	61 71	61 30	49 51	68 03
98 57	96 85	89 19	98 86	98 40	71 02	74 14	84 01	74 86	16 19	18 14	66 33
28 89	91 08	01 01	07 20	63 21	51 82	15 85	02 90	50 62	01 32	96 20	49 26
25 14	19 22	29 84	92 87	78 82	29 91	39 48	33 13	14 18	30 46	30 46	35 28
93 46	89 39	25 21	43 05	41 83	92 46	22 88	69 97	16 24	25 95	97 95	16 72
32 23	63 73	41 78	94 76	33 88	98 89	47 70	53 68	90 76	31 81	91 48	20 67
77 66	93 02	21 84	58 73	01 72	62 17	78 76	42 40	93 63	42 56	87 17	12 59
03 85	71 71	34 17	98 19	15 64	21 34	29 86	44 56	11 00	22 16	99 40	39 04
50 01	30 84	13 25	36 74	11 55	56 34	07 15	67 15	10 67	95 13	11 95	27 33

08 39	47 88	24 05	54 12	98 67	08 83	37 40	06 53	06 22	89 56	74 41	47 05
72 09	29 63	56 92	29 42	95 24	86 01	63 57	26 82	37 85	92 31	89 71	41 50
32 31	86 79	75 09	72 37	16 81	94 32	50 03	97 43	41 71	81 61	90 35	79 54
43 00	18 60	85 43	65 59	69 44	89 85	66 30	49 45	73 49	38 00	42 03	36 86
30 72	70 68	37 51	60 08	01 02	04 80	05 26	53 84	58 64	66 46	23 18	05 14
73 47	62 29	40 94	14 03	24 60	51 10	67 60	22 06	48 06	32 77	79 95	72 65
42 95	29 60	55 04	70 74	19 86	38 79	30 99	82 64	59 55	92 99	56 64	55 78
82 28	21 85	36 74	34 08	93 42	91 17	50 41	87 60	74 06	20 60	30 80	89 17
82 06	98 34	19 16	05 08	43 53	71 69	22 42	88 61	85 13	94 95	45 77	21 32
10 26	34 28	18 80	25 54	72 93	53 20	80 94	26 92	01 11	23 36	02 27	45 86

72 48	75 56	92 71	52 88	65 22	12 11	26 97	54 77	85 69	74 03	05 81	93 60
43 57	65 26	90 94	29 97	83 56	16 80	35 73	29 98	29 61	81 65	60 20	78 74
93 74	09 87	16 21	89 64	82 95	55 39	78 77	14 16	09 63	55 03	84 88	85 14
69 21	79 24	11 02	76 37	92 23	00 92	07 59	86 92	48 16	05 13	94 49	10 97
74 54	52 05	21 04	38 02	28 01	85 14	33 59	45 24	07 56	45 72	22 37	99 26
01 92	17 82	44 88	38 72	81 89	85 60	56 74	02 16	28 07	55 83	09 70	41 47
17 88	24 71	57 24	49 74	43 16	46 90	68 97	01 71	91 57	01 76	19 70	87 06
69 46	81 37	41 63	63 06	84 93	24 11	93 80	02 00	87 12	34 33	90 88	61 42
40 06	60 93	47 65	32 88	29 44	83 12	69 87	90 09	55 22	23 00	60 42	59 26
18 92	34 43	51 20	67 25	18 15	07 91	54 41	84 90	65 86	39 20	57 32	50 42

61 31	38 06	49 73	63 30	56 71	46 78	21 33	16 40	61 80	40 80	35 57	65 46
60 84	04 59	77 17	70 99	18 66	59 27	92 13	26 88	18 97	92 46	01 18	78 89
50 92	61 66	62 17	30 36	24 71	78 59	89 26	07 36	40 54	72 13	21 45	73 16
37 51	08 23	67 82	38 24	84 65	70 83	63 02	22 66	69 19	86 40	87 60	05 39
99 05	75 78	42 83	23 80	13 71	58 99	53 25	89 33	17 74	93 39	78 88	81 84
23 11	78 92	94 15	61 12	86 92	10 94	55 52	22 44	56 32	71 22	33 34	69 18
07 41	55 76	61 40	12 38	23 19	58 35	98 24	30 79	76 45	22 59	70 58	50 72
18 76	63 42	69 17	20 39	94 71	70 54	88 19	84 48	14 51	24 65	52 61	92 60
09 35	01 25	89 06	69 42	32 52	39 79	09 89	10 74	43 24	35 97	37 79	45 96
74 34	36 99	08 73	15 61	33 24	02 80	29 28	62 94	13 32	27 45	15 24	96 35

B2

5	46 87
5	24 91
1	10 94
1	59 07
0	00 13
3	51 88
2	14 50
9	86 04
3	62 51
9	22 89
47	54 00
12	77 56
86	80 13
39	39 39
90	35 22
74	24 47
00	06 83
93	28 58
58	13 66
60	35 68
19	60 63
10	53 45
69	19 32
66	62 77
02	02 55
44	42 72
02	79 32
83	13 40
47	40 17
22	35 13
37	93 60
9 51	68 03
8 14	66 33
6 20	49 26
0 46	35 28
7 95	16 72
1 48	20 67
7 17	12 59
9 40	39 04
1 95	27 33
74 41	47 05
39 71	41 50
90 35	79 54
42 03	36 86
23 18	05 14
79 95	72 65
56 64	55 78
30 80	89 17
45 77	21 32
02 27	45 86
05 81	93 60
60 20	78 74
84 88	85 14
94 49	10 97
22 37	99 26
09 70	41 47
19 70	87 06
90 88	61 42
60 42	59 26
57 32	50 42
35 57	65 46
01 18	78 89
21 45	73 16
87 60	05 39
78 88	81 84
33 34	69 18
70 58	50 72
52 61	92 60
37 79	45 96
15 24	96 35

TABLE 6:

RANDOM NUMBERS

C3

96 73	55 04	92 01	85 20	71 33	81 57	24 65	69 55	06 52	43 15	09 23	09 83
25 44	72 88	39 34	61 77	69 55	63 60	23 76	26 89	83 90	26 57	44 77	68 02
35 12	90 56	12 44	30 65	67 85	19 12	16 73	69 60	72 81	55 31	19 32	13 49
01 65	01 18	25 33	98 51	70 02	84 00	32 26	87 63	95 80	90 53	14 03	05 14
91 71	20 61	96 19	82 43	17 40	92 84	45 84	76 87	89 38	36 47	29 13	28 43
71 99	34 05	01 91	32 67	84 84	54 08	26 96	44 03	13 23	72 49	84 96	92 54
96 40	41 90	95 90	58 43	45 94	13 23	44 28	90 56	20 96	62 12	91 04	93 69
51 46	17 95	24 76	66 05	72 62	97 34	81 32	45 25	69 58	52 58	46 64	96 87
31 72	57 39	05 03	30 89	89 10	78 35	83 69	41 52	67 77	03 81	58 17	28 23
41 53	05 92	89 94	80 90	00 64	29 48	18 44	67 05	25 91	50 02	63 11	83 23
60 41	08 41	20 06	17 88	74 96	27 65	81 63	08 45	23 90	03 60	86 20	12 45
25 78	51 26	36 87	75 58	33 39	96 94	71 62	48 08	96 47	08 69	64 91	25 36
70 82	11 97	55 77	56 69	39 91	16 37	37 00	59 61	04 35	68 57	65 85	04 80
66 80	47 97	20 55	75 03	95 63	08 77	09 78	49 69	57 60	25 97	35 90	79 52
69 25	34 09	94 50	17 88	44 85	65 34	69 20	15 94	46 53	38 26	59 32	90 23
43 41	23 40	50 40	80 21	58 51	89 20	06 54	56 28	22 47	62 26	65 54	73 41
53 01	02 42	78 55	35 83	47 23	44 08	72 53	88 54	91 42	13 12	07 59	89 68
94 96	43 01	32 56	61 52	14 93	55 75	02 42	26 89	07 44	89 14	31 31	52 63
40 72	21 73	17 78	52 40	10 57	37 39	50 03	00 81	01 19	09 90	93 43	12 64
35 11	03 59	73 25	32 56	35 14	60 23	25 08	56 48	42 52	70 30	91 37	76 18
32 27	87 24	65 03	37 20	91 94	20 83	52 34	72 40	74 28	10 00	78 69	65 18
28 95	53 82	90 65	98 68	19 92	56 47	14 30	44 60	49 77	89 46	96 42	81 15
09 50	24 33	28 04	36 98	42 68	38 60	52 08	34 37	02 04	04 33	42 13	93 44
85 99	99 96	27 42	61 42	38 02	11 47	33 40	93 50	50 84	78 81	35 94	86 00
49 20	10 12	71 17	30 84	08 16	90 35	41 59	85 11	73 63	27 48	09 15	60 18
98 89	35 95	71 22	49 79	78 36	86 83	49 78	62 08	94 47	53 61	66 03	71 42
26 39	77 36	92 76	76 58	38 68	98 78	70 97	47 22	21 88	15 25	55 61	07 68
94 24	14 89	67 21	39 19	23 49	05 57	26 92	95 33	48 75	28 20	72 11	57 11
28 15	03 34	70 83	56 04	08 47	54 38	41 13	98 40	94 15	88 49	09 17	84 15
06 54	63 45	15 25	59 22	33 21	61 14	86 82	49 44	80 64	33 98	62 28	28 17
09 71	12 62	73 42	71 53	18 59	40 51	01 95	17 55	52 45	36 85	33 43	40 40
42 92	90 41	42 72	44 18	68 65	48 57	06 88	37 93	32 75	88 13	00 82	24 50
25 99	64 83	09 99	30 80	92 79	62 07	05 50	22 43	35 38	39 39	50 25	04 73
18 18	67 84	42 26	28 21	32 81	54 36	34 58	21 24	36 98	27 11	63 89	70 76
56 84	09 26	40 56	21 75	04 87	08 06	45 32	72 98	92 17	47 51	28 59	02 04
52 19	60 56	06 97	95 47	40 26	95 24	97 74	34 93	62 82	01 19	80 07	38 89
69 54	37 27	31 17	37 83	40 68	88 88	76 21	67 16	11 52	41 73	85 35	23 33
34 63	73 57	80 85	56 36	66 00	43 86	43 24	42 43	40 21	81 45	53 24	41 44
14 46	08 16	74 38	67 13	17 26	12 27	63 33	43 42	91 72	65 78	78 27	48 27
54 63	26 19	55 10	27 33	43 98	30 00	26 04	90 35	57 56	93 64	25 22	47 99
04 08	92 53	12 21	35 21	34 62	67 74	54 06	97 46	88 03	44 05	01 89	23 33
69 51	91 92	60 52	03 08	25 48	28 29	46 82	12 65	18 13	91 60	39 55	50 55
31 27	91 93	97 28	62 73	81 76	18 99	10 72	35 53	52 96	94 86	71 57	07 05
68 76	92 88	15 93	18 20	62 51	81 90	48 80	93 53	19 17	77 63	33 29	18 42
55 63	61 91	49 99	56 01	82 13	53 87	58 45	84 39	81 01	57 86	58 97	09 90
53 28	38 42	55 95	18 75	60 79	91 61	50 76	92 27	56 18	71 97	14 30	64 26
94 19	68 85	81 72	62 78	11 28	86 38	90 65	82 84	75 37	29 40	68 09	45 63
76 18	52 12	39 26	03 57	23 70	68 94	13 96	59 18	85 69	03 70	90 62	60 37
98 87	20 78	35 23	30 23	76 45	55 64	43 62	55 24	20 28	56 76	69 77	82 98
61 24	84 03	53 90	22 26	41 19	12 99	05 64	71 47	00 68	95 65	60 39	31 35
11 96	28 09	72 26	67 50	03 15	28 83	36 04	33 19	30 85	20 69	14 05	02 01
88 43	22 55	24 41	51 22	56 86	79 61	42 40	59 18	15 10	29 72	70 79	19 60
19 96	90 68	52 91	32 85	19 62	29 39	49 47	01 53	01 62	29 76	89 67	10 58
35 79	42 30	71 55	21 15	82 78	21 95	27 98	97 20	19 95	28 46	25 34	02 88
72 34	73 14	35 96	21 38	65 19	44 47	23 63	09 65	89 20	93 17	57 31	48 86
73 17	88 62	48 08	78 34	13 83	66 23	26 38	98 10	07 78	57 23	66 52	07 38
74 10	29 42	43 35	35 42	65 98	09 77	89 12	58 87	48 73	50 22	18 13	15 58
65 23	14 19	00 88	96 29	07 00	46 06	10 26	14 48	27 17	24 15	69 13	14 56
49 17	50 50	65 54	21 98	31 66	37 96	01 64	94 76	44 02	72 99	62 63	02 08
77 40	20 88	88 65	04 89	52 65	69 37	11 64	90 78	55 27	59 26	43 59	36 87
73 71	44 50	93 65	01 62	11 67	65 84	70 77	45 78	21 99	71 43	00 85	37 70
09 66	12 69	46 65	84 54	12 09	38 20	29 60	92 86	77 48	94 74	31 88	65 85
86 09	87 48	34 00	20 93	01 16	57 51	85 81	20 43	23 99	67 01	74 55	45 00
65 64	59 13	51 84	97 10	38 70	37 30	46 01	39 37	89 77	97 59	81 98	69 98
55 11	33 17	00 37	80 90	65 41	53 65	30 96	11 85	95 52	40 53	67 57	51 13
51 10	16 80	68 60	15 03	52 67	25 52	21 90	74 16	29 14	18 66	38 21	17 74
24 00	98 62	77 78	26 31	51 77	29 06	76 74	11 30	63 98	82 82	35 41	07 12
19 97	37 56	77 19	62 81	11 90	52 18	60 57	89 33	06 22	47 65	15 99	55 49
82 82	12 03	88 85	82 78	33 14	91 36	57 15	20 12	96 24	03 59	38 97	78 77
82 05	04 57	11 25	87 92	95 30	84 50	14 02	97 33	83 96	90 67	72 82	98 89

APPENDIX - MATHEMATICAL PRINCIPLES IN DISEASE SURVEYS

A.1 Testing for the Presence of a Disease

Suppose that we are sampling a population to test that there are no positives in the population.

Let

- N be the population size,
- d be the number of diseased animals in the population
- n be the number of animals sampled

and β be the probability that there are no positives in the sample.

Either from first principles, or from the hypergeometric distribution, these four variables are related by

$$\beta = \frac{N-d}{N} \cdot \frac{N-d+1}{N-1} \cdots \frac{N-d-n+1}{N-n+1} \qquad (1)$$

which may be rewritten as

$$\beta = \frac{(N-d)!\ (N-n)!}{N!\ (N-d-n)!} \qquad (2)$$

Equation (1) simply states that the probability that there are no positives in the sample is equal to the probability a positive is not the first animal sampled times the probability that a positive is not the second animal sampled, and so on, to the probability that a positive is not the n^{th} animal sampled.

In order to answer the question 'How many to sample?', we take the N and d of interest and successively multiply the terms in (1) until we get a product less than the desired level. If we choose this value of n as our sample size, the chances of finding at least one of the d positives will be $1-\beta$, the confidence level that we want.

It is more tedious to answer the question 'If I find no positives in a sample what is an upper limit to the number of positives?' Simplistically trial and error could be used to solve (1) for d given N, n and β. Luckily there is a very good approximation available. Each term in the numerator and denominator of (1) can be approximated by the 'middle' term of the product to give:

$$\beta \simeq \left(\frac{N - d - (n-1)/2)}{N - (n-1)/2}\right)^n = \left(1 - \frac{d}{N - (n-1)/2}\right)^n$$

By taking the n^{th} root, we find that d is given by:

$$d \simeq (1 - \beta^{1/n})(N - \frac{n-1}{2})$$

and this is basically the approximation referred to in Table 1.

probability no positives in the sample

If we look at equation (2) we see that d and n can be interchanged without changing the formula. This is why Table 1 can be used to answer both questions. It also means that the same form of approximation can be used to quickly obtain the sample size:

$$n \approx (1 - \beta^{1/d})(N - \frac{d-1}{2})$$

As the population size increases but $\theta = d/N$, the proportion of diseased animals, remains constant, then equation (1) becomes

$$\theta = (1-\beta)^n \qquad (3)$$

(This is equivalent to the binominal approximation to the hypergeometric distribution.)

Solving (3) for n gives:

$$n = \log \beta / \log(1-\theta)$$

as the limit to the sample size required, and

$$\theta = 1 - \beta^{1/n}$$

as the upper limit to the proportion of diseased animals in the population if no positives are found in a sample from a very large population.

A.2 Estimating a Proportion:

This section considers the problems concerning the estimation of the actual level of disease, rather than simply determining its absence or presence.

Again we shall consider a population of size N with $d = N\theta$ diseased animals. We shall take a sample of size n and look at the (random) number X of diseased animals in the sample.

The statistical distribution that describes this form of sampling is the hypergeometric distribution. From simple combinational arguments we can show that the probability that we get x positives in the sample is

$$\text{Prob}\ \{X=x\} = \binom{d}{x} \binom{N-d}{n-x} \quad \binom{N}{n}$$

The mean of X is $n\theta$ and the variance is $n\theta(1-\theta)\frac{N-n}{N-1}$.

If the population size is large, the binomial distribution can be used to approximate the hypergeometric distribution to give

$$\text{Prob}\ \{X=x\} = \binom{n}{x} \theta^{x} (1-\theta)^{n-x}$$

X has mean and variance of $n\theta$ and $n\theta(1-\theta)$ respectively.

We want to answer three questions:

i) what is an estimate of θ?

ii) what are confidence limits for θ?

iii) how big a sample should we take to get specified confidence limits? (which should have been asked first).

i) Estimating the proportion

Regardless of the population size, we shall use $p=x/n$ as an estimate of the proportion of diseased animals.

ii) Finding confidence limits for a proportion

Initially we shall assume that the sample is from an infinite population and later give an approximation to take into account the population size.

When determining confidence limits, we are looking for values of θ for which the observed value is 'likely'. Put another way, if the probability of observing the same or worse result for a particular value of θ is small, then that value of θ is not included in the confidence limits. This reasoning leads to solving the equation

$$\beta = \sum_{i=0}^{x} \binom{n}{i} \theta^{i} (1-\theta)^{n-i} \qquad \text{for } \theta.$$

There is no elegant easy alternative to using a systematic trial and error approach.

Table 5 provides these confidence limits.

The effect of the finite population size is to reduce the variance by the fraction (N-n)/(N-1). We should then expect that the width of the confidence limits would be reduced by a factor of

$$\sqrt{\frac{N-n}{N-1}} \text{ or approximately } \sqrt{(1-\frac{n}{N})}.$$

This approximation turns out to be quite good for middle sized values of θ, but it over estimates the reduction for lower limits near $\theta = 0$, and upper limits near $\theta = 1$ (because the variable cannot go outside the range of 0 and 1).

For large values of n (say greater than 50), the normal approximation to the binomial distribution holds fairly well at the tails of the distribution. If v is the corresponding normal percentage point, then approximate confidence limits are given by

$$p \pm \left(v \sqrt{\frac{p(1-p)}{n}\left(1-\frac{n}{N}\right)} + \frac{1}{2n}\right)$$

iii) <u>Approximate sample size required</u>

Before doing the sampling, it is worthwhile estimating the sample size required to give the desired width of confidence limits, say $\pm w$.

Using the normal approximation mentioned above, this requirement means that, for an infinite population,

$$w = v\sqrt{\theta(1-\theta)/n}$$

By solving this for n, we obtain

$$n_{\infty} = (v/w)^2\,\theta(1-\theta)$$

and it is this function that is given in Table 4.

For a finite population, the requirement implies that

$$w = v\sqrt{\frac{\theta(1-\theta)}{n}\left(1-\frac{n}{N}\right)} = v\sqrt{\theta(1-\theta)\left(\frac{1}{n}-\frac{1}{N}\right)}$$

By squaring and rearranging we get

$$\frac{1}{n} = \frac{1}{N} + \left(\frac{w}{v}\right)^2 \frac{1}{\theta(1-\theta)} = \frac{1}{N} + \frac{1}{n_{\infty}}$$

A.3 Estimating Population Size

Sometimes we do not know the population size, but rather are wanting to estimate it. Typical cases would concern feral animals or where mustering is difficult.

One method is to use a capture-recapture sampling scheme.

Suppose that an initial sample of d animals are captured, marked in some way and released. After a time suitable to allow for mixing of the population, but which would preclude many deaths/births, another sample of n (the recapture) is done. X is the number of the original sample recaptured.

An estimation of the population size is given by $N = d\frac{n}{x}$

Approximate confidence limits can be found (using Table 5) as follows

(i) estimate $N = d\frac{n}{x}$
(ii) calculate $p = d/N$, $q=1-p$, $f=n/N$
(iii) calculate upper and lower limits for p (from Table 5) after adjusting the width by the finite population correction factor $(\sqrt{1-f})$
(iv) estimate upper and lower estimates using d/p.

Often only an upper limit for N is required. In this case only the lower one-sided limit for p needs to be calculated.

Example 1:

At the first round of testing of a herd, 120 animals were mustered. At the second round, 130 animals, which included 110 of those present at the first test, were mustered. How many animals are untested?

We have d=120, n=130 and x=110

(i) This gives an estimate of $N=120 \times 130 \div 110 = 141$

(ii) $p=.85$, $q=.15$, $f=.92$, $\sqrt{1-f}=.28$

(iii) from the 95% table the lower limit is
$.85 - (.07 \times .28) = .85 - 0.0196 = .8304$

(iv) whence 95% confidence limit for the maximum value of N is 144

Thus: 110 were tested twice
10 were presented for first test but not second
20 were presented for second test but not first
4 remain untested.

Example 2:

400 feral pigs are captured, marked and then released. 40 of the original capture are found when another 400 pigs are captured.

We have d=400, n=400, and x=40.

(i) N=4000
(ii) $p=.1$, $q=.9$, $f=.9$, $\sqrt{1-f} = .95$
(iii) upper limit is $.1 + (.025 \times .95) = .1037$
lower limit is $.1 - (.025 \times .95) = .0763$
(iv) 95% limits for N are (3233, 5249).

A.4 Further Reading

Cochran, W.G. (1963), Sampling Techniques. Second ed. John Wiley and Sons Inc., New York.

Cochrane, A.L. and Holland, W.W. (1971) Validation of screening procedures. Br. Med. Bull. 27 (1):3-8.

Leech, F.B. and Sellers, K.C. (1979). Statistical Epidemiology in Veterinary Science. Charles Griffin and Co. Ltd, London.

Slonin, M.J. (1967). Sampling: A quick reliable guide to practical statistics. Simon and Schuster, New York.

Stewart, G.T. (1970) Epidemiological approach to assessment of Health. Lancet 2, 18 July: 115-119.

Trajstman, A.C. (1979). Diagnostic tests: sensitivity, specificity, efficacy and prevalence. Aust. vet. J. 55(10): 501.